FLICC / CENDI Symposium, Dec. 11 2001

OAIS: From Requirements to Reality at OCLC

Pam Kircher

Product Manager, Digital Archive

OCLC Digital & Preservation Resources

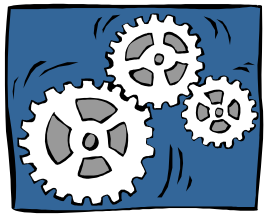
OCLC and FirstSearch are registered trademarks of OCLC Online Computer Library Center,

CORC is a trademark of OCLC Online Computer Library Center,

OCLC Digital Archive

Long-term retention and access

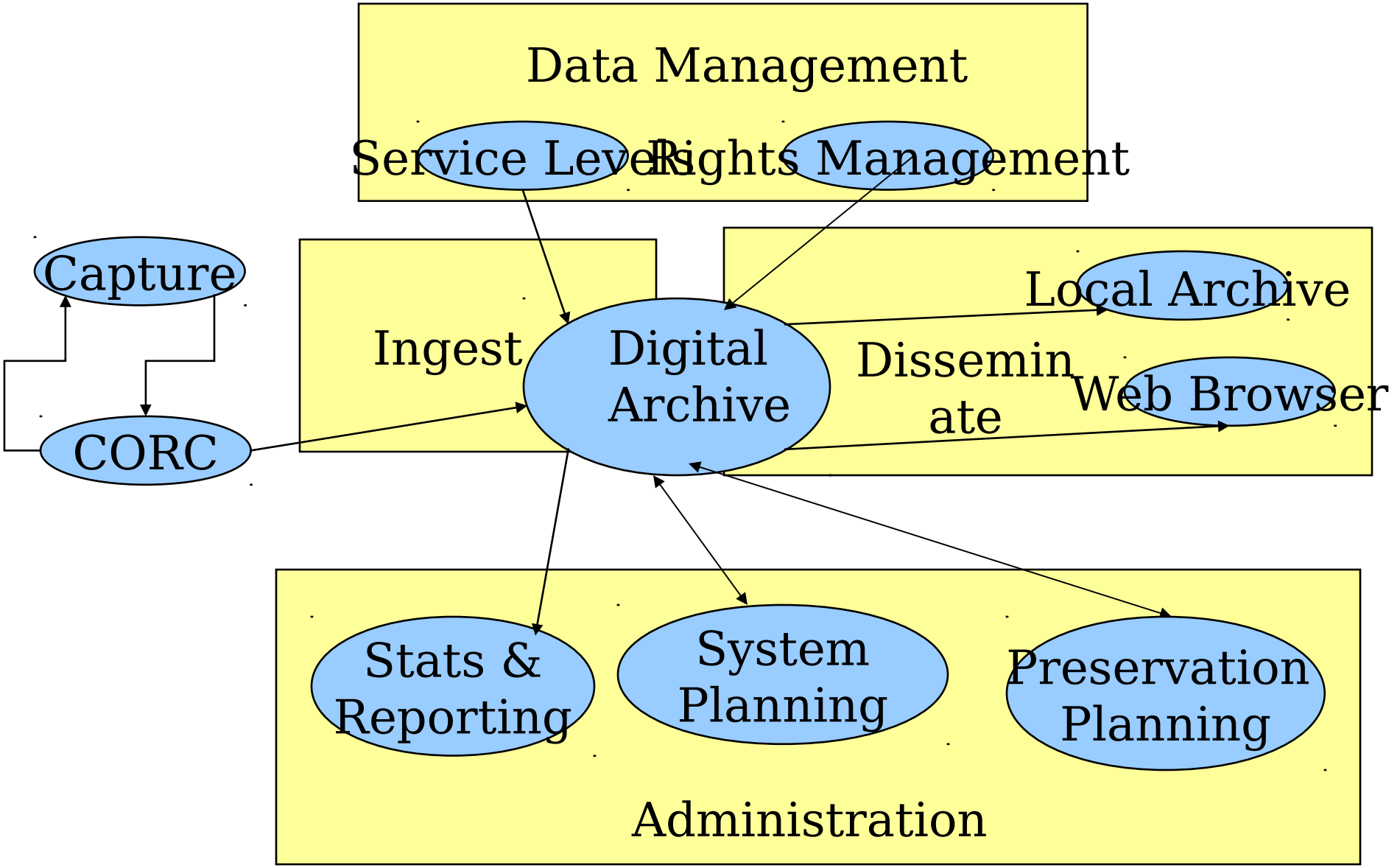
- Interoperable
  - OAIS
  - Preservation Metadata



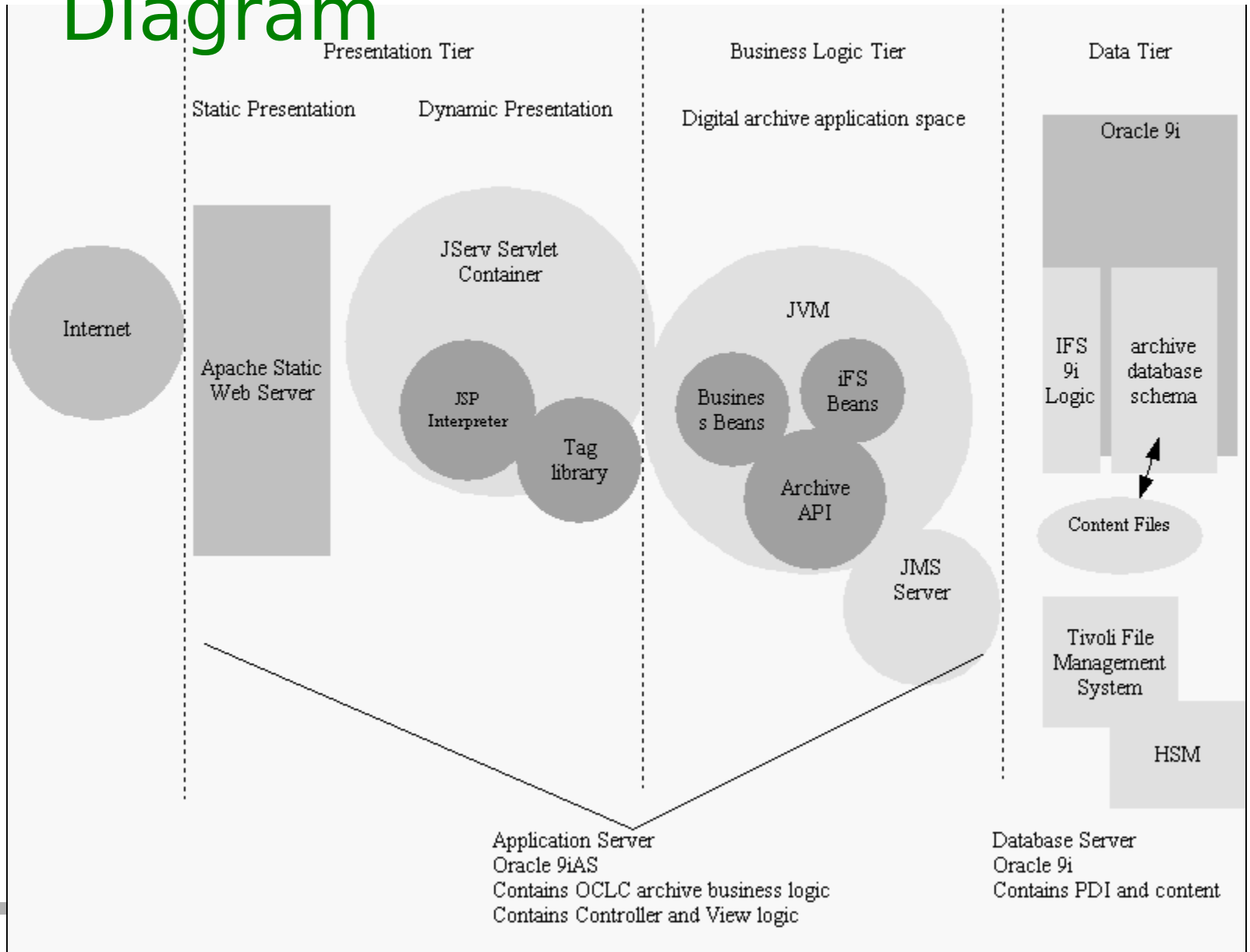
- Choice of service levels
- Integrate with current workflo

  - CORC-based tools
  - Administration module

OAIS to OCLC Digital Archive



OCLC Digital Archive Tier Diagram



All Digital Archive Service Levels

- OCLC admin staff:
  - Performance and media management
  - Periodic QA for functionality & fixity
  - Offsite backup
- Owner admin staff:
  - Movement of objects from one service level to another
  - Content management

Digital Archive Service Levels

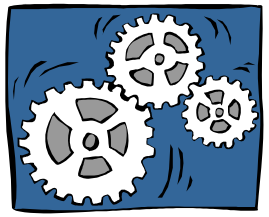
Service Level	Store	Access	Preserve
Tools Only			
Basic Backup	✓	Dark	
Long-term Preservation	✓	Dark	✓
Active Access	✓	Active	
Active Archive	✓	Active	✓

Implementation Drivers at OCLC

- Object characteristics
  - Born digital
  - Web documents
  - Mostly public-domain
- User characteristics
  - Didn’t create the object
  - Want to integrate workflows
  - Use current staff
- Supporting tools
  - CORC
  - Content and Autho Groups

Web Document Digital Archive Pilot

- Implement digital archive
- Manage web-based documents
  - Capture
  - Long-term retention & access
- Develop best-practices
  - Preservation metadata
  - Workflows
- Direct input from users



OCLC Web Document Digital Archive

Browser or OPAC

INTERNET

FirstSearch

- Search WC
- View Object

CORC

- Bib metadata
- Pres metadata

Harvester

- Web Crawl
- Crawl profile
- Capture
- Manual review

Digital Archive

Authentication

•

Ingest/validate

- Admin interface

•

Dissemination

- Storage
- Retrieval

… other repositories

OCLC Digital Archive Record

- Based on OAIS information model
- 28 elements plus sub-elements
  - Descriptive, preservation, representation
  - Still images and text
- Implemented in XML
- Evolving

Capture

- User directed harvesting interface
  - Preview
  - Review
- Virus checking and checksum

Ingest

- Representation information
  - Structure of web document
- Packaging information

Dissemination

- Objects and metadata (DIP) via FTP
- View
  - Via standard browsers
  - PURL/URL syntax is OpenURL
  - Administrator sets access rights
  - Administrator creates collections

Next phases

- Batch ingest
- Migration, on-the-fly conversion & emulation
- PURL re-direct
- Capture improvements
- Digital rights management
- Document authenticity issues
- More file types

Questions?

www.oclc.org

pam_kircher@oclc.org